DOCUMENT RESUME

ED 144 394 FL 008 897

AUTHOR Oller, John W., Jr.
TITLE How Important is Language Proficiency to IQ and Other Educational Tests? Occasional Papers on Linguistics, No. 1.
INSTITUTION Southern Illinois Univ., Carbondale. Dept. of Linguistics.
PUB DATE Apr 77
NOTE 12p.; Proceedings of the International Conference on Frontiers in Language Proficiency and Dominance Testing (1st, Southern Illinois University, Carbondale, Illinois, April 21-23, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Cognitive Processes; *Educational Testing; Intelligence Quotient; *Intelligence Tests; *Language Proficiency; Language Research; Language Skills; Language Tests; Language Usage; Neurological Organization; Psycholinguistics; Psychological Tests; Test Reliability; *Test Validity

ABSTRACT

It is often the case that intelligence tests do not measure what they purport to measure, although it is commonly assumed that every test with a different name is a test of different skills, and that tests with the same name test the same skills. Spearman attempted to study the relationship between tests that supposedly measured different mental abilities, and upon finding substantial correlation between different tests, posited a general factor of intelligence ("g"), to explain these relationships. The present study questions this approach. If "g" turned out to be indistinguishable from language proficiency, for example, questions would arise as to the implications for intelligence testing and the use of IQ scores, and for school tests that claim to measure constructs other than language proficiency but rely heavily on language-based performances. These questions are considered from three angles: (1) the pattern of interrelationships among tests that claim to measure different aspects of intelligence and language tests that require discourse processing is examined; (2) intelligence tests themselves are examined to analytically define what they require of learners; and (3) evidence from brain physiology and the processing of information is considered. (CLK)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

How Important is Language Proficiency to IQ and Other Educational Tests?

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

John W. Oller, Jr.

Southern Illinois University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Southern Illinois University

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED144394

In logic it is usually necessary to distinguish the symbol from what it symbolizes. Similarly, in educational testing it makes sense to distinguish tests from whatever they purport to measure. But, as we will see, such distinctions are not always made, and the failure to make them inevitably leads to perplexing problems. The point is nicely illustrated by the following story about Abraham Lincoln. It seems that he was involved in an argument with a particulary stubborn opponent when he hit upon a stroke of homespun genius reminiscent of the tale about the blind men and the elephant.

"How many legs does a cow have?" he asked rhetorically.

"Four, of course," his antagonist replied impatiently.

Lincoln nodded, "That's right. Now suppose we call the cow's tail a leg, how many legs would the cow have?"

"Why, five, of course," his opponent replied without hesitation.

"Now that's where you're wrong," Mr. Lincoln answered, "simply calling a cow's tail a leg doesn't make it a leg" (Brown and Carlsen, 1955, p.6).

Cronbach was reported to have said that the most common error in tests and measurement is assuming that every test with a different name is a test of a different skill, and that all tests with the same name are therefore tests of the same skill. But calling a test by a certain name does not guarantee that it is a test of whatever the name purports to refer to. Just because the Stanford-Binet happens to be the granddaddy of all so-called "IQ" tests doesn't necessarily make it a test of intelligence, and the fact that the Brown-Carlsen is called a test of listening comprehension doesn't necessarily preclude the possibility that it may be a test of intelligence or something else.

Perhaps no one would expect educators, psychometrists, or people who use tests to make the error of assuming that a test is a measure of whatever its name implies, but the error is in fact a common one. It is not difficult to find examples. For instance, the renowned Berkeley psychologist, Arthur Jensen (1972) asserts that "intelligence is what intelligence tests measure" (p. 8). He likens it to "electricity" which he asserts "is easier to measure than to define" (p. 5). Now here is where the logical problem arises. Is it intelligence or measures of intelligence that we are considering? Is electricity

FL008897

nothing more than what measures of electricity measure? If this were so, common folk would have to redefine what they perceive as lightening as something related to measures of electricity--and nothing more. But common people assume that electricity has some existence independent of measures of it. If their assumption were not correct, it would be necessary to say that electricity was invented when the first measure of electricity was devised and that before it could be measured, it didn't exist.

If intelligence is nothing more than what intelligence tests measure, it isn't a bit like electricity, because we see and feel and experience the benefits of electricity quite apart from measures of it. We know, for instance, what it is to experience an electric shock, to see a spark, to hear the thunder that lightening creates, to use an electric appliance. Isn't the same true of intelligence in the common sense of the term? Don't we know what it is to have a burst of insight, to learn, to discover, to know, to express, to communicate, to think, to plan, to evaluate? Further, we are reluctant to attribute any of these activities to things that lack intelligence--such as rocks and petrified trees.

The following interview with an ape that graduated from Harvard illustrates the point. It appeared in the February 4 issue of the Southern Illinois University, Daily Egyptian:

Language Needs Monkeying

Several people have asked what that Harvard-educated gorilla plans to do now that he's lost his bid for the White House. Actually, he's devoting all his energies to a cause dear to his heart--The Society for the Prevention of Condescension to Animals.

The Ivy League intellectual, who has a vocabulary of 500 words said he became interested in the SPCA the day he learned his 500th word. It was "chairperson".

"I have never heard a more offensive word in my life," said the bare knuckled fighter angrily. "I can put up with racists, sexists, and ageists, but words like these are clearly promulgated by humanists.

"What, I ask you, is wrong with 'chaircreature'? ...Phoney liberals are always telling me that some of their best friends are beasts," he said. "And I even met a lady once who claimed she had married one. She was mistaken, but when I pointed out her error," said the gentle hearted idealist, his sensitive lower lip quivering, "she took me aside and told me straight out that 'beast' and 'beastly' refer to that which is 'disgusting' or 'abominable'! I've never been so hurt."

The fierce battler for social justice demanded that all human chauvinist people stop referring to "beasts" unless they follow the word with the phrase "of the field" which he said created a "lovely pastoral image".

But to stamp out subliminal humanism once and for all, the dedicated amateur lexicographer said it would be necessary to create substitutes for a number of other offensive words including

"brute, "monkey around", and "swine flu".

"I'm sure there are thousands of others," he apologized, but altogether I only know 500."

The report of an ape graduating from Harvard merely confirms certain widespread rumors, but the idea that he could have learned enough there to have a conversation about it is surprising.

Clearly, there is an aspect of intelligence in the use of language that we regard as distinctively human. Fromkin and Rodman (1974) point out that in a certain African culture, a child is not considered to have transcended the level of mere things until he begins to speak. Then he is considered to have become human. Apart from what Chomsky (1972) referred to as "some interesting science fiction" about dolphins, there is no evidence that organisms other than human beings are capable of the multiple degrees of abstraction that even a child can attain in talking about who said what and how so-and-so responded and why, etcetera. While all non-human organisms are not dumb, and some of them may seem more humane than humans on occasion, they simply cannot approximate the abstract cognitive performances of a normal twelve year-old--much less a political candidate making a bid for the White House. Hence, whatever intelligence is, it surely involves and is involved in what people do with language.

What then is intelligence? Jensen (1972) warns against asking this question, because he contends that it has no answer (p.6). Yet he himself cannot resist quoting the speculations of others on the "question of what intelligence <u>really</u> is" nor can he refrain from offering certain views of his own. Although he contends that "the trouble comes only when we attribute more to 'intelligence' and to our measurements of it than do the psychologists who use the concept in its proper sense" (1972, p.8), the reader is left in somewhat of a quandary concerning just what that "proper sense" might be. He says, "the best we can do is to obtain measurements of certain kinds of behavior and look at their relationships to other phenomena and see if these relationships make any kind of sense and order. It is from these orderly relationships that we can gain some understanding of the phonemena" (p.6).

What Jensen is leading up to is a discussion of the fact that superficially different tests may show deep patterns of interrelationship. Near the turn of the century, Charles Spearman developed the now widely used statistical technique of factor analysis. He was interested in studying the relationship between different tests that were supposed to be measures of different mental abilities. By his methods, disparate tests were shown to be often substantially correlated--that is, they produced overlapping variances. For instance, a task requiring the manipulation of objects in various spatio-temporal relationships may be shown to correlate substantially with what is called a test of vocabulary. To explain these sorts of relationships, Spearman posited a general

factor of intelligence which he abbreviated "g". Of this factor, Jensen says, "Despite numerous theoretical attacks on Spearman's basic notion of a general factor, g has stood like a Rock of Gibraltar in psychometrics, defying any attempt to construct a test of complex problem solving which excludes it" (p.9).

What then is intelligence really? Jensen explains by example: "Standard intelligence scales such as the Binet and the Wechsler are composed of a dozen or so subtests which differ obviously in their superficial appearance: vocabulary, general information, memory span for digits, block designs, figure copying, mazes, form boards, and so on. When the intercorrelations among a dozen or more such tests are subjected to a factor analysis or principal components analysis, some 50 percent or more of the total individual differences variance in all the tests is usually found to be attributable to a general factor common to all the tests. Thus, when we speak of intelligence it is this general factor, rather than any single test, that we should keep in mind" (p. 10).

But what is the general factor really? Jensen quotes Spearman's speculation that all of the tests that load on (i.e., correlate with and thus contribute to the definition of) g "seemed to involve 'the ability to educe relations and correlates', " which Jensen elaborates on as the ability "to see the general from the particular and the particular as an instance of the general" (p.9). He continues to quote Aquinas who saw intelligence as the ability "to combine and separate" and again he elaborates: "To see the difference between things which seem similar and to see the similarities between things which seem different. These are essentially the processes of abstraction and conceptualization. Tasks which call for problem solving requiring these processes are usually the best measures of g" (p. 9).

Interestingly, the processes and tasks that Jensen regards as being most characteristic of intelligence are precisely the sorts of things what are also the most characteristic of the use and learning of language. Psychologists such as George Miller and Charles Osgood have long emphasized the importance to language of precisely the sorts of processes referred to by Jensen as being at the heart of the definition of intelligence. In fact, the principles of induction and substitution have been proposed as the basis of language use as well as language learning (Oller, 1972). It is not difficult to show that these same principles are the same as those referred to by Aquinas, Spearman, and Jensen in their characterizations of intelligence.

Consider for instance, the kind of mental operations involved in what Jensen calls "cross modal transfer"--a process which he regards as epitomizing the notion of intelligence. A person may be shown a bunch of wooden blocks and then after being blindfolded he may be asked to identify the blocks by sense of touch. Or, a child may be asked to name the letter that an adult writes with his finger on the child's back. As Jensen notes, almost anyone can

perform these intelligence tests. They demonstrate the fact that human beings by virtue of their intelligence can easily cope with entirely novel relationships between what may be quite novel stimulus patterns in different sensory modalities. Is it not also true that every normal use of language for communication involves an exceedingly complex form of "cross modal transfer"? Hence, is it not also true that language ability is at the hub of whatever the wheel of intelligence may be?

If we merely follow the trail that Jensen has marked off, we are led to the conclusion that the g factor may well be coextensive with the ability to use a language--with what has been variously called language competence, communicative competence, a conceptual dependency grammar, an expectancy system, and so forth. What if the widely acknowledged factor of intelligence--that is, g--were really indistinguishable from language proficiency? What would the implications for intelligence testing and the use of so-called "IQ" scores be? Further, what would the implications be for a wide variety of school tests that purport to measure constructs other than language proficiency, but which rely heavily on language based performances?

There are other lines of reasoning that also lead us to the conclusion that what intelligence tests measure may be a lot more like language proficiency than it is like electricity. Even if we follow the suggestions of proponents of the idea that intelligence is something more mysterious and less definable than language proficiency, we are still led to the speculation that the algebraically definable variance produced by tests that are called tests of intelligence may be indistinguishable from similar quantities produced by discourse processing tasks such as cloze procedure and dictation.

Empirical tests can be applied to the question from at least three different angles. First, we may examine the pattern of interrelationships among tests that purport to measure different aspects of intelligence (e.g., "verbal" and "non-verbal IQ" tests) and language tests that require discourse processing; second, we may look at the tests themselves and attempt analytically to see what they require learners to do; and third, we may consider evidence from studies of the physiology of the brain itself and the ways that it processes information.

The first method is essentially an application of Spearman's factoring technique. It is the one recommended by Jensen and is applied by several of the contributors to this volume. The crucial questions for this approach pertain to the observed correlations between tests that purport to measure the same thing in contrast to tests that purport to measure different things. For instance, do tests that aim at "verbal intelligence" correlate more strongly with each other than with tests that aim at, say, "reading ability"? Or do tests that aim at overall "intelligence" correlate more substantially with each other than with tests that purport to measure

global language proficiency? There are at least as many ways of posing these sorts of empirical questions as there are of naming a cow's legs and tail. The crucial thing to bear in mind is that the names do not necessarily change the nature of the objects named.

The research on the topic of what different tests actually measure is extensive in spite of the fact that it is relatively sparse in view of the vast number of tests. We will concentrate here on a few tests that have been widely used. To begin with we will consider the claims for a test that purports to measure non-verbal intelligence--Raven's Progressive Matrices. This test was inspired by the Binet test which is better known in its revised form as the Stanford-Binet. What the test purports to measure is best told by its own author:

> The Colored Matrices are designed for use with young children and old people, for anthropological studies and for clinical work. They can be used satisfactorily with people who for any reason cannot understand the English language, with people suffering from physical disabilities, aphasias, cerebral palsy, or deafness, as well as with people who are intellectually subnormal or have deteriorated. The scale as a whole can be described as "a test of observation and clear thinking". Each problem in the scale is really the "mother" or source of a system of thought, while the order in which problems are presented provides the standard training in the method of working. . . . By itself, it is not a test of "general intelligence", and it is always a mistake to describe it as such. For this purpose, it should be used in conjunction with a vocabulary test (Raven, 1956, p. 3).

Unfortunately, the experts do not all agree on what the matrices measure or do not measure. Jensen sees them and Cattell's Culture Fair Tests as "tests . . . which measure g in nearly pure form" (1972, p.9).

Since the experts do not agree it would seem all the more important to find out what in fact the Raven test is a test of. Observed correlations reported in the Manual range from .65 to .73 with various tests of vocabulary and are not any higher for tests that purport to measure non-verbal IQ. Hence, Jensen's claim that the matrices measure g seems to be sustained.

A second source of information of an empirical sort is to examine the content of the tests themselves. An application of this technique to a wide range of tests is offered by Gunnarson (this volume). If we apply the same technique to the Raven Matrices we discover that they consist of objects or patterns that must be placed in certain relationships to each other in order to complete a visual "matrix". Supposedly the ability

to complete such patterns or matrices is indicative of the subject's capacity to grasp abstract relationships. For example, consider Figure 1 which is an item from one of Raven's matrices.

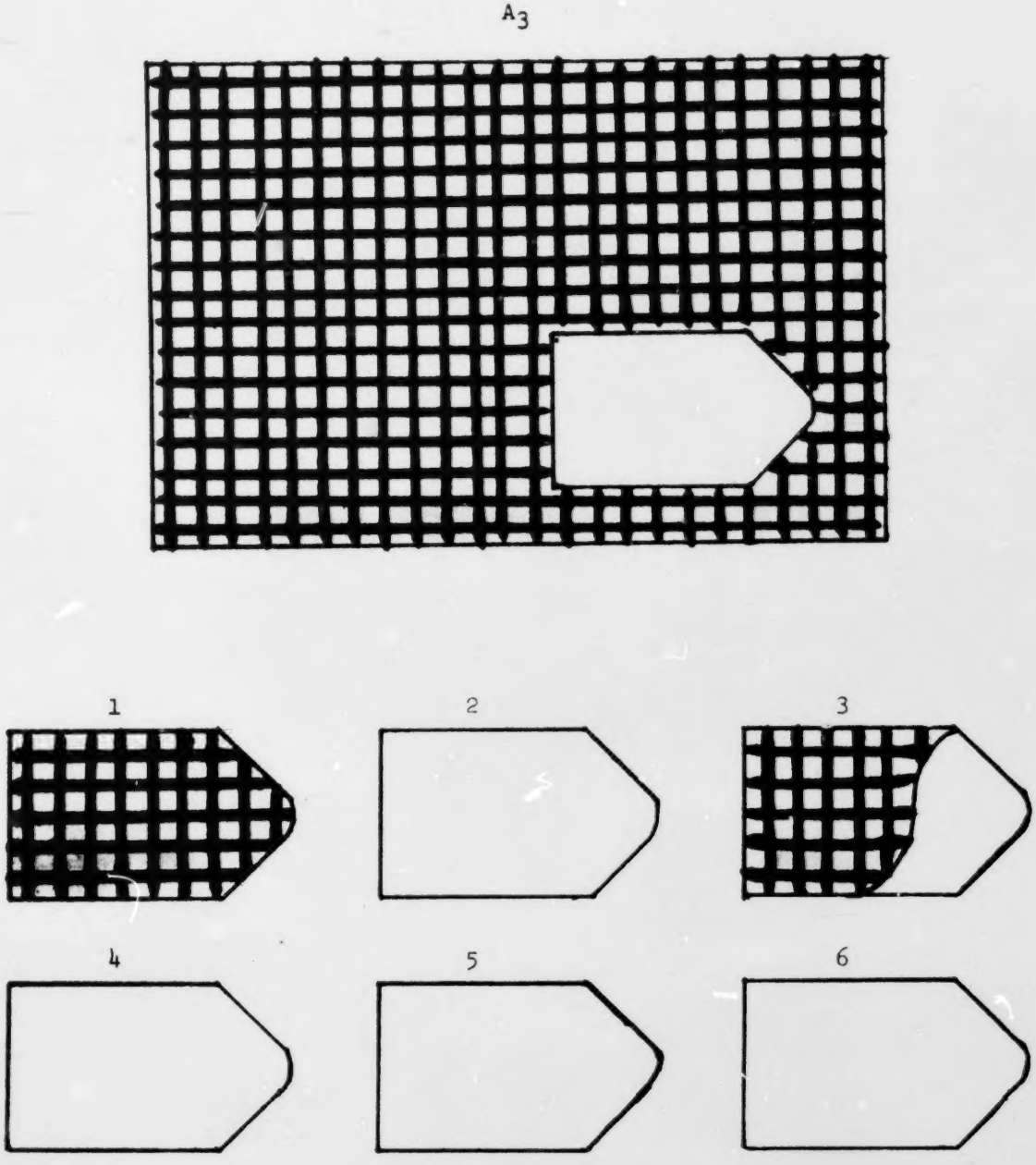

Figure 1. A Sample item from the Raven Progressive Matrices, a "non-verbal IQ" test.

Further, consider the following verbal instructions and ask yourself whether the test would be suitable if the directions were addressed to you in a language you had limited knowledge of, say, Mandarin, or Objibwa: "Look at this," the psychologist says, "a bit of this pattern has been cut out. Which of these is the right one to go in there? This one fits in, but it is not right. That is not right either. That is right, isn't it? That is wrong, isn't it? It is wrong here. You put the right one in. That is right, shall we do another one? We have done that one right, let us put it at the side here. . . . See if you can find the bit that goes in here. . . . Now you see how to do them, put this one over here. . . . Take the next one and see if you can do it yourself. . . . Here are some more, look at the figures in each column. Which piece comes from here. Notice how the figures go along each row. What comes here? Point to the right one" (from p. 15 of the Guide).

Is it quite unreasonable to suggest that there is an element of language comprehension in such a task? It is just possible that the rather strong correlations previously observed between the Raven test and so-called verbal IQ tests would remain even if the verbal instructions could be gotten out of the Raven test, but it is not likely that such verbal instructions can ever be done without completely. Jensen (1972) notes that the Stanford-Binet served as the basic model for the Raven test, and McNemar (1942) found that a single common factor could explain essentially all of the variance in the latter test. Flahive (this volume) explores correlations between the Raven test and E.S.L. proficiency and reading scores. His findings suggest that Raven's test is rather strongly correlated with language tests.

Following the same line of reasoning concerning the relation between language tasks and so-called IQ tests, we can easily extend the argument to encompass other educational tests. Table 1 displays correlations between the California Test of Mental Maturity (an IQ test), the Terman McNemar Test of Mental Ability, the Nelson Denny Reading Test, and the Kelley Greene Reading Comprehension Test with the Brown-Carlsen Listening Comprehension Test.

Table 1

Correlations between the Brown-Carlsen Listening Comprehension Test and Other Educational Measures (as reported in the Manual).

Other Tests	The Brown-Carlsen LC
California Test of Mental Maturity	.78, .67
Terman McNemar Test of Mental Ability	.70, .74
Nelson Denny Reading Test Total	.63
Kelley Greene Reading Comprehension Test	.65, .66

It is curious that the so-called "listening" test correlates more strongly with the "IQ" tests than it does with the "reading" tests.

Table 2 presents similar data correlating the widly used School and College Ability Tests (regarded as IQ measures) with the Sequential Tests of Educational Progress (supposedly measures of achievement). However, the average correlations based on as many as 26 different test applications with subject populations of 200 to 225 between total scores on the two tests are typically above .70. The implication is that within the limits of error of measurement, the tests may be largely measures of the same thing. Since the STEP tests are avowedly oriented toward language skills

Table 2

Correlations between the School and College Ability Tests (SCAT) and the Sequential Tests of Educational Progress (STEP), from ETS (1958).

STEP Subtests	SCAT Verbal	SCAT Quantitative	Correlations for Total Scores
Reading*	.77 to .84	.55 to .74	.71 to .85
Writing*	.70 to .83	.51 to .69	.66 to .79
Listening**	.65 to .78	.49 to .67	.66 to .79

(namely, "Reading", "Writing", and "Listening"), we are again led to ask whether IQ tests might not be measures of language proficiency as much or perhaps more than anything else. Stump (this volume) pushes the question further by investigating dictation and cloze scores attained by native speakers of English in relation to scores on IQ and achievement tests.

Another source of information, in addition to content analysis and statistical analyses of test variances, is the investigation of brain functions. It is interesting that certain language functions of the brain have been rather narrowly localized according to Morgan (1965) as shown in Figure 2. This figure was displayed in a mimeographed pamphlet distributed by Slosson Educational Publications as part of an advertisement or propaganda sheet recommending the use of various Slosson tests.

BRAIN FUNCTIONS AND DYSFUNCTIONS

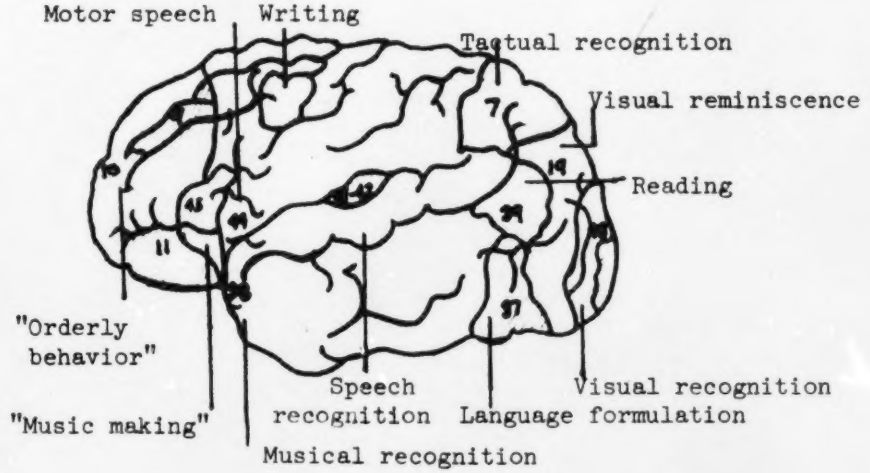

Figure 2.

It is depressing that the tenor of the brochure is so thoroughly addressed to failure and deviancy. A few remarks from the pamphlet will illustrate the tone:

> Brain dysfunction, malformation, impairment, damage or injury often go undiagnosed as the difficulty "does not show" and may not be too evident. Persons with such a "hidden" handicap are often awkward and show poor "eye-hand" coordination. Many of them have difficulty in drawing the various figures on the Slosson Drawing Coordination Test (SDCT), which supplements (SIT). They may have poor memory and other "weak areas" of mental ability, showing wide "scatter" on the Slossen Intelligence Test (SIT) for Children and Adults. They may read poorly on the Slosson Oral Reading Test (SORT).
>
> Children or adults who show sufficient difficulty on the SDCT, SIT, or SORT may require careful neurological, psychological, and psychiatric evaluations in order that their particular weakness be diagnosed and dealt with wisely and intelligently. . . .
>
> The SDCT, the SIT, and the SORT are offered as testing instruments helpful in the early screening and identification of some of these handicapped individuals and are intended for use by professional persons who deal with problems of education, guidance, and counseling (no date, p.1).

What proof is there that the diagnoses that can be made with the kinds of tests referred to are sound? What is the evidence from neurolinguistics to support the claims for separate brain functions that pertain to presumed components of language skill? For lack of space it is not possible to do a thorough review of the research, but the conclusions of someone who has done such a review may be cited. Krashen (1973) concludes that perhaps

> there are, properly speaking, no language areas of the brain: there are only cortical mechanisms that are useful in the production and comprehension of language. The mechanisms responsible for temporal processing and the use of transformational relationships should perhaps no more be thought of a primary language mechanisms than the lungs, the larynx, the palate, the nose, the tongue, the teeth, and the lips be considered the primary organs of speech. Physiologically, language may be an overlaid function (p. 53).

Ultimately, the findings of psycholinguistic, sociolinguistic, and neurolinguistic research concerning the nature of linguistic and other sorts of cognitive and affective processes must be

reconciled. Furthermore, it seems logical to suppose that the outcomes of research in the various disciplines must ultimately lead to similar conclusions regarding the nature of human intelligence and the relation of intelligence to the use and learning of language. In the meantime, it seems reasonable to push ahead with testing educational tests in order to try to determine by empirical methods what in fact such tests are tests of. At the present, the accumulating research evidence seems to suggest that a vast array of educational tests that go by many different names may be measures of language proficiency more than of anything else. The most important implication for education is related to the fact that most people are willing to assume that language proficiency can be modified by instructional procedures, or by mere experience in communicating in the language, while the same people are probably not willing to make a similar assumption about "IQ".

"Well then," Mr. Lincoln might have said to the high school counselor, or the college dean, "can language proficiency be modified?"

"But of course," the educators would undoubtedly have answered.

"Suppose we call the child's language proficiency his IQ," Mr. Lincoln might have continued, "could the child's intelligence change as a result of experience in and out of the classroom?"

"Of course, his innate IQ couldn't change very much as a result of experience," the unwary educator might reply.

"Now that's where you're wrong," Mr. Lincoln would assert, "simply calling language proficiency IQ doesn't make it equivalent with innate intelligence."